Problem Sets

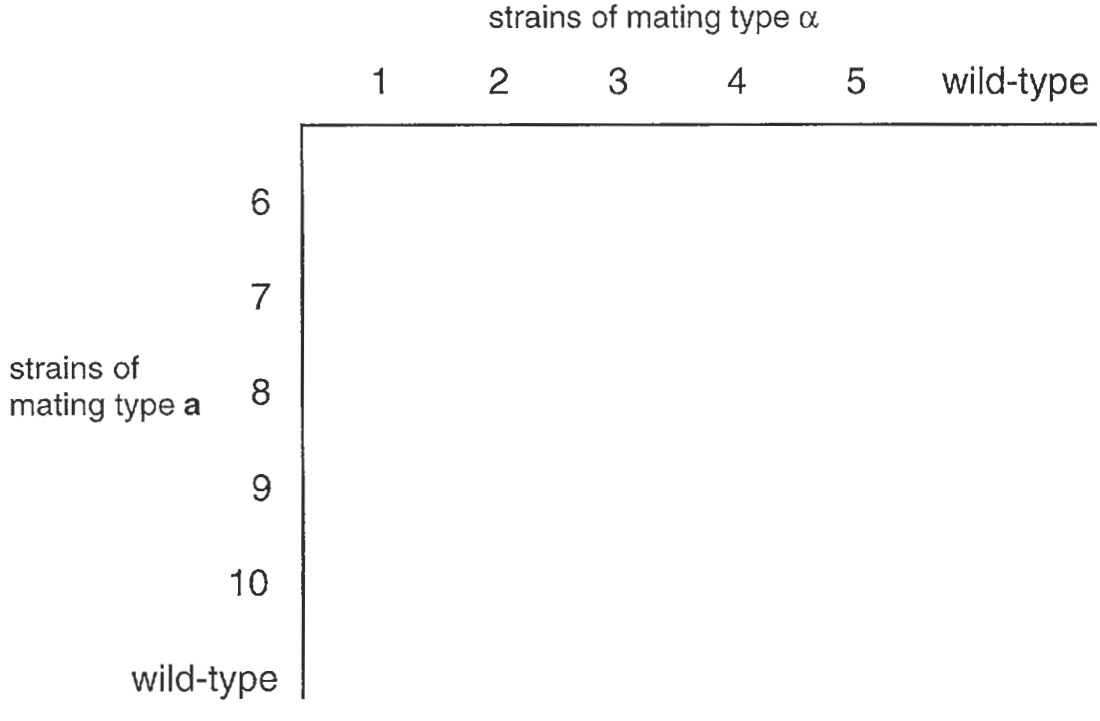
Fall 1997

7.03 Problem Set 1

due before 5 PM on Thursday, September 18

Hand in answers in recitation section or in the box outside the class

1. (a) In class, we discussed how yeast auxotrophic mutants can be used for genetic experiments. Mutants that give resistance to toxic compounds are also useful for genetic analysis because resistance can be easily scored in tests for growth on plates containing the toxic compound. Canavanine is an analog of the amino acid arginine that is highly toxic to yeast cells. Let's say that you are sent a set of mutants that are resistant to canavanine (numbered 1 – 10). Mutants 1 – 5 are of mating type α and mutants 6 – 10 are of mating type **a**. Say that mutations in three different genes can give resistance to canavanine. Strains 1 and 8 have recessive mutations in the first gene, strains 2, 3, and 7 have recessive mutations in the second gene, and strains 4, 9, and 10 have recessive mutations in the third gene. Strains 5 and 6 have dominant alleles that confer canavanine resistance. In order to verify the genotypes of the strains you have received, you set up all possible crosses among the strains including crosses to wild-type. Reproduce the table below and indicate a "+" at the intersection of the two parental strains if you expect the diploid that results from the cross to be resistant to canavanine and indicate a "–" where you expect the diploid to be sensitive to canavanine.



(b) You are interested in genes involved in histidine synthesis and have isolated four new His^- mutants in the **a** mating type. We will call these mutants 1, 2, 3, and 4. You cross each mutant to each of the following test strains: MATα His^+, MATα $His2^-$, and MATα $His3^-$. The results of the crosses are given in the table below where a "+" indicates growth of the diploid on medium that lacks histidine and a "–" indicates no growth of the diploid on medium that lacks histidine.

	Mutant 1	Mutant 2	Mutant 3	Mutant 4
MATα His^+	+	–	+	+
MATα $His2^-$	+	–	–	+
MATα $His3^-$	–	–	–	+

Give as complete a description as you can of each of the new mutants based on the results of these tests. Be sure to note any ambiguities that remain.

2. Although it is always advisable to use true-breeding lines for genetic experiments there are some traits for which true-breeding strains cannot be constructed. A classic example is the yellow coat color allele in mice. Mice that are heterozygous for this allele have yellow fur, but mice that are homozygous are dead.

(a) Is the yellow allele dominant or recessive for yellow fur? Is the yellow allele dominant or recessive for lethality?

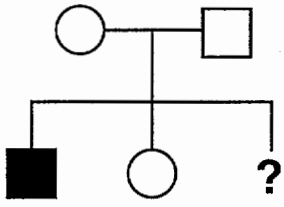
(b) Given a yellow mouse and a wild-type mouse how would you show that the allele for the yellow trait is dominant. Describe the cross(es) that you would do and explain how the results would be interpreted.

(c) In a cross between a yellow mouse and a wild-type mouse what is the probability of no yellow mice in a litter of six? In a cross between two yellow mice what is the probability of no yellow mice in a litter of six?

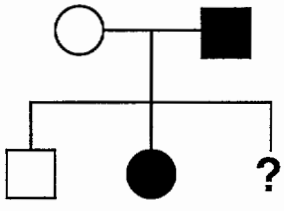
(d) A cross between two yellow mice produces a litter of ten with seven yellow mice and three wild-type mice. In test crosses of the yellow F_1 mice to wild-type, some wild-type progeny are produced in each test cross. Use the Chi-square test to determine whether the distribution of genotypes observed among the litter of ten is significantly different ($p < 0.05$) from the distribution expected if the yellow allele were not lethal in homozygotes. Please give the observed and expected genotypic ratios, the degrees of freedom, and your calculated value for $\chi 2$.

3. For each of the following pedigrees the individuals expressing a trait are indicated by solid symbols. Consider the following possible modes of inheritance (i) autosomal recessive, (ii) autosomal dominant, and (iii) X-linked recessive. For each pedigree state which, if any, of the modes of inheritance are not possible unless either the occurrence of a new mutation or incomplete penetrance is invoked. For the modes of inheritance that are possible, calculate the probabilities that the next child indicated by a (?) will be affected. In the case of X-linked recessive inheritance, calculate separate probabilities for sons and daughters.

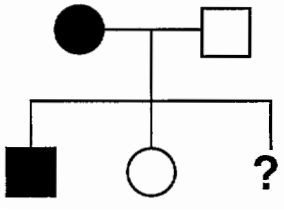
(a)



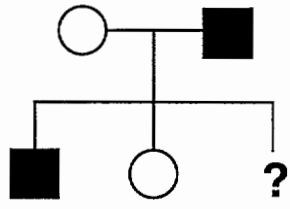
(b)



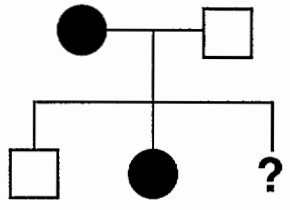
(c)



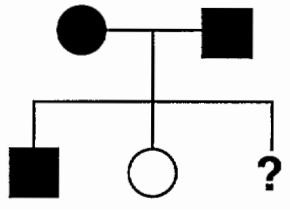
(d)



(e)



(f)



7.03 Problem Set 1
Solutions

due before 5 PM on Thursday, September 18
Hand in answers in recitation section or in the box outside the class

1. (a) In class, we discussed how yeast auxotrophic mutants can be used for genetic experiments. Mutants that give resistance to toxic compounds are also useful for genetic analysis because resistance can be easily scored in tests for growth on plates containing the toxic compound. Canavanine is an analog of the amino acid arginine that is highly toxic to yeast cells. Let's say that you are sent a set of mutants that are resistant to canavanine (numbered 1 – 10). Mutants 1 – 5 are of mating type α and mutants 6 – 10 are of mating type **a**. Say that mutations in three different genes can give resistance to canavanine. Strains 1 and 8 have recessive mutations in the first gene, strains 2, 3, and 7 have recessive mutations in the second gene, and strains 4, 9, and 10 have recessive mutations in the third gene. Strains 5 and 6 have dominant alleles that confer canavanine resistance. In order to verify the genotypes of the strains you have received, you set up all possible crosses among the strains including crosses to wild-type. Reproduce the table below and indicate a "+" at the intersection of the two parental strains if you expect the diploid that results from the cross to be resistant to canavanine and indicate a "–" where you expect the diploid to be sensitive to canavanine.

		strains of mating type α					
		1	2	3	4	5	wild-type
strains of mating type **a**	6	+	+	+	+	+	+
	7	–	+	+	–	+	–
	8	+	–	–	–	+	–
	9	–	–	–	+	+	–
	10	–	–	–	+	+	–
	wild-type	–	–	–	–	+	–

1a) Because strains 5 and 6 have dominant alleles that confer canavanine resistance, all crosses involving them will result in a canavanine-resistant diploid strain . Since all the other strains have recessive mutant alleles, all crosses with wild-type will result in a diploid strain without canavanine resistance.

Because strains 1 and 8 are in the same complementation group, crosses between strains 1 and 8 will result in a homozygous mutant diploid; this diploid will therefore be resistant. Similarly, because strains 2, 3, and 7 are in a second complementation group, crosses involving any two strains within that group will result in a homozygous mutant diploid resistant to canavanine. The same situation is true for crosses within strains 4, 9, and 10.

If a particular mutant strain is crossed to different strain which has a mutation in a different complementation group, the resulting diploid will not be resistant to canavanine; complementation has occurred.

(b) You are interested in genes involved in histidine synthesis and have isolated four new His^- mutants in the **a** mating type. We will call these mutants 1, 2, 3, and 4. You cross each mutant to each of the following test strains: MATα His^+, MATα $His2^-$, and MATα $His3^-$. The results of the crosses are given in the table below where a "+" indicates growth of the diploid on medium that lacks histidine and a "–" indicates no growth of the diploid on medium that lacks histidine.

	Mutant 1	Mutant 2	Mutant 3	Mutant 4
MATα His^+	+	–	+	+
MATα $His2^-$	+	–	–	+
MATα $His3^-$	–	–	–	+

Give as complete a description as you can of each of the new mutants based on the results of these tests. Be sure to note any ambiguities that remain.

1b) Mutation 1 is a recessive mutation, in the same complementation group as the $His3^-$ mutation. Mutation 1 is probably an allele of HIS3.

Mutation 2 is dominant to wild-type. No informantion about complementation can be determined for this mutation.

Mutation 3 is a recessive mutation. It fails to complement both the $His2^-$ and $His3^-$ mutations. Since HIS2 and HIS3 are different genes, Mutation 3 could either be a double mutation in both the HIS2 and HIS3 genes, or a large deletion that covers both loci.

Mutation 4 is a recessive mutation that complements both the $His2^-$ and $His3^-$ mutations. Mutation 4 could be a mutation in a novel gene required in the histidine biosynthesis pathway.

2. Although it is always advisable to use true-breeding lines for genetic experiments there are some traits for which true-breeding strains cannot be constructed. A classic example is the yellow coat color allele in mice. Mice that are heterozygous for this allele have yellow fur, but mice that are homozygous are dead.

(a) Is the yellow allele dominant or recessive for yellow fur? Is the yellow allele dominant or recessive for lethality?

2a) The yellow allele is dominant for yellow fur, but recessive for lethality.

(b) Given a yellow mouse and a wild-type mouse how would you show that the allele for the yellow trait is dominant. Describe the cross(es) that you would do and explain how the results would be interpreted.

2b) Cross the yellow mouse with the wild-type mouse, and look at the coat color of the F1 progeny. If all the F1 progeny had wild-type coat color, then the allele for yellow coat color is recessive. If half of the F1 progeny had wild-type coat color, with the other half having yellow coat color, then the allele for yellow coat color is dominant.

(c) In a cross between a yellow mouse and a wild-type mouse what is the probability of no yellow mice in a litter of six? In a cross between two yellow mice what is the probability of no yellow mice in a litter of six?

2c) Each newborn mouse in a litter is considered an independent event. In a cross between a yellow mouse and a wild-type mouse, the probability of a single F1 progeny being non-yellow is 1/2. Hence, the probability of a litter of 6 having no yellow mouse is $(1/2)^6 = 0.016$ In a cross between two yellow mice, the probability of an F1 progeny being yellow is 2/3, and the probability of being non-yellow is 1/3. Therefore, the probability for a litter of 6 having no yellow pups is $(1/3)^6 = 0.0014$

(d) A cross between two yellow mice produces a litter of ten with seven yellow mice and three wild-type mice. In test crosses of the yellow F_1 mice to wild-type, some wild-type progeny are produced in each test cross. Use the Chi-square test to determine whether the distribution of genotypes observed among the litter of ten is significantly different ($p < 0.05$) from the distribution expected if the yellow allele were not lethal in homozygotes. Please give the observed and expected genotypic ratios, the degrees of freedom, and your calculated value for $\chi 2$.

2d) The hypothesis is : the yellow allele does not result in lethality in homozygotes.

	Observed	*Expected*
yy :	*0*	*2.5*
y+ :	*7*	*5*
++ :	*3*	*2.5*

$\chi^2 = \Sigma (observed - expected)^2 / expected = 3.4$

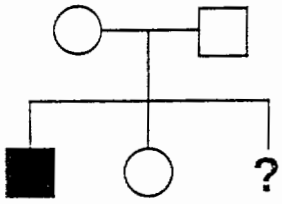
degrees of freedom = (number of possible events) - 1 = 3 - 1 = 2

For $\chi^2 = 3.4$ ====> $0.5 > p > 0.1$

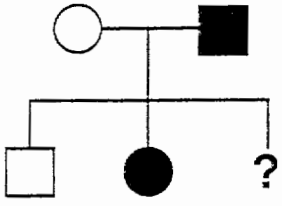
This p value indicates that the observed value does not deviate significantly from the expected values when assuming that the yellow allele does not result in lethality in homozygotes.

3. For each of the following pedigrees the individuals expressing a trait are indicated by solid symbols. Consider the following possible modes of inheritance (i) autosomal recessive, (ii) autosomal dominant, and (iii) X-linked recessive. For each pedigree state which, if any, of the modes of inheritance are not possible unless either the occurrence of a new mutation or incomplete penetrance is invoked. For the modes of inheritance that are possible, calculate the probabilities that the next child indicated by a (?) will be affected. In the case of X-linked recessive inheritance, calculate separate probabilities for sons and daughters.

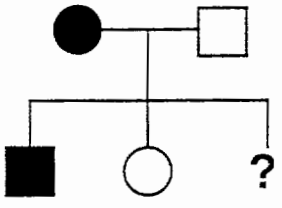
(a)



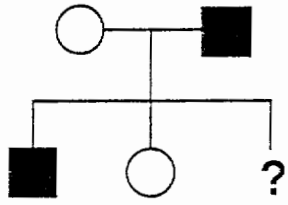
(b)



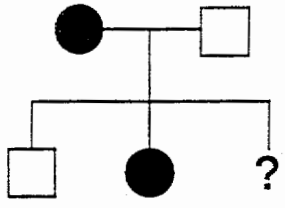
(c)



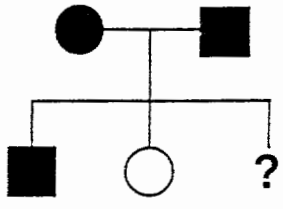
(d)



(e)



(f)



Inheritance pattern	*Parental Genotypes*	*Probability*
a) Autosomal recessive	*(♀)Aa x (♂)Aa*	*p(?) = 1/4*
X-linked recessive	*(♀)X^aX x (♂)XY*	*p(?) = 0, if female* *p(?) = 1/2 if male*
b) Autosomal recessive	*(♀)Aa x (♂)aa*	*p(?) = ~~3/4~~ 1/2*
Autosomal dominant	*(♀)aa x (♂)Aa*	*p(?) = 1/2*
X-linked recessive	*(♀)X^aX x (♂)X^aY*	*p(?) = 1/2 if female* *p(?) = 1/2 if male*

Problem 3 solutions, continued...

Inheritance pattern	*Parental Genotypes*	*Probability*
c) Autosomal recessive :	(♀)*aa x* (♂)*Aa*	*p(?) =* 1/2
Autosomal dominant :	(♀)*Aa x* (♂)*aa*	*p(?) = 1/2*
X-linked recessive :	(♀)X^aX^a *x XY*	*p(?) = 0 if female* *p(?) = 1 if male*
d) Autosomal recessive	(♀)*Aa x* (♂)*aa*	*p(?) = 1/2*
Autosomal dominant :	(♀)*aa x* (♂)*Aa*	*p(?) = 1/2*
X-linked recessive	(♀)X^aX *x* (♂)X^aY	*p(?) = 1/2 if female* *p(?) = 1/2 if male*
e) Autosomal recessive	(♀)*aa x* (♂)*Aa*	*p(?) = 1/2*
Autosomal dominant	(♀)*Aa x* (♂)*aa*	*p(?) = 1/2*
f) Autosomal dominant	(♀)*Aa x* (♂)*Aa*	*p(?) = 3/4*

7.03 Problem Set 2

due before 5 PM on Thursday, September 25

Hand in answers in recitation section or in the box outside the class

1. Suppose that there are two different mouse genes, 40 cM apart, that when mutated will produce mice with six toes. Further, suppose that the combination of mutations in both genes is lethal. You are given two different six-toed mice each from a different true-breeding line. A cross of the two mice produce F_1 mice which all look normal.

(a) What can you say about the two true-breeding lines that produced the mice that you were given?

You now inter-cross F_1 mice to produce F_2 progeny.

(b) What fraction of the F_2 mice that are born will look like wild-type? Remember you must take into account both recombination and the lethality of double mutants. This problem should make it clear why mapping studies are, whenever possible, done by crossing heterozygous individuals to a homozygous recessive test strain.

2. Consider the three hypothetical recessive traits in Drosophilia: bent wings (bt), yellow body (yl), and long bristles (lg). You are given a fly that looks like wild-type, but is heterozygous for all three traits. You set up a test cross of your to a that is homozygous recessive for all three traits. From 500 progeny the following types are found (abbreviations are used for recessive traits and the corresponding dominant traits are indicated by a +).

Phenotype	number
bt yl lg	34
+ + +	46
bt + +	13
+ yl lg	7
bt yl +	202
+ + lg	193
bt + lg	3
+ yl +	2

(a) If the fly that you were given came from a cross of two true-breeding lines, what would be the phenotypes of the two true-breeding lines?

(b) Produce a genetic map showing the order and the relative positions of the genes that control the three traits.

3. You have isolated three new yeast mutants that are copper-resistant . Mutant 1 is dominant for resistance and is in mating type α. Mutants 2 and 3 are recessive and are in mating type **a**. You mate mutant 1 to mutant 2, sporulate the resulting diploid and dissect tetrads. Out of 25 tetrads, 4 have two copper resistant spores and two copper sensitive (wild-type) spores, 5 have four copper-resistant spores, and 16 have three copper-resistant spores and one wild-type spore.

(a) What can you conclude about the relative positions of mutation 1 and mutation 2 ?

(b) If instead of dissecting tetrads you had examined 100 random spores, what fraction of the spores would you expect to be copper-resistant? What fraction of the copper-resistant spores would you expect to produce copper-resistant diploids when mated to wild-type?

Next, you mate mutant 1 to mutant 3, sporulate the diploid and dissect tetrads. Of 25 tetrads, 21 have four copper-resistant spores, and 4 have three copper-resistant spores and one wild-type spore.

(c) What can you conclude about the relative positions of mutation 1 and mutation 3 ?

(d) If instead of dissecting tetrads you had examined 100 random spores, what fraction of the spores would you expect to be copper-resistant? What fraction of the copper-resistant spores would you expect to produce copper-resistant diploids when mated to wild-type?

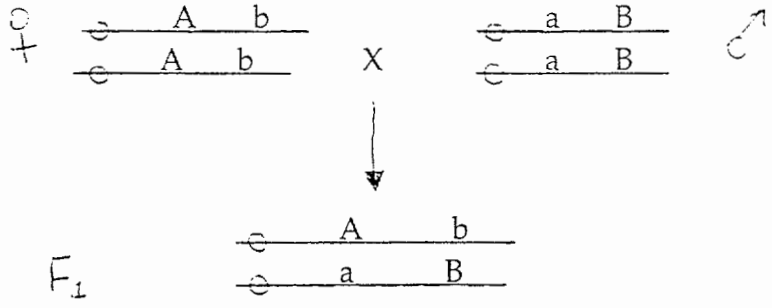
7.03 Problem Set 2 ANSWERS

1. Suppose that there are two different mouse genes, 40 cM apart, that when mutated will produce mice with six toes. Further, suppose that the combination of mutations in both genes is lethal. You are given two different six-toed mice each from a different true-breeding line. A cross of the two mice produce F_1 mice which all look normal.

(a) What can you say about the two true-breeding lines that produced the mice that you were given?

Because all of the F1 progeny are phenotypically normal, you know that the mutations in each of the parental strains complement each other. Therefore, each true-breeding line is homozygous for a recessive mutation in different genes.

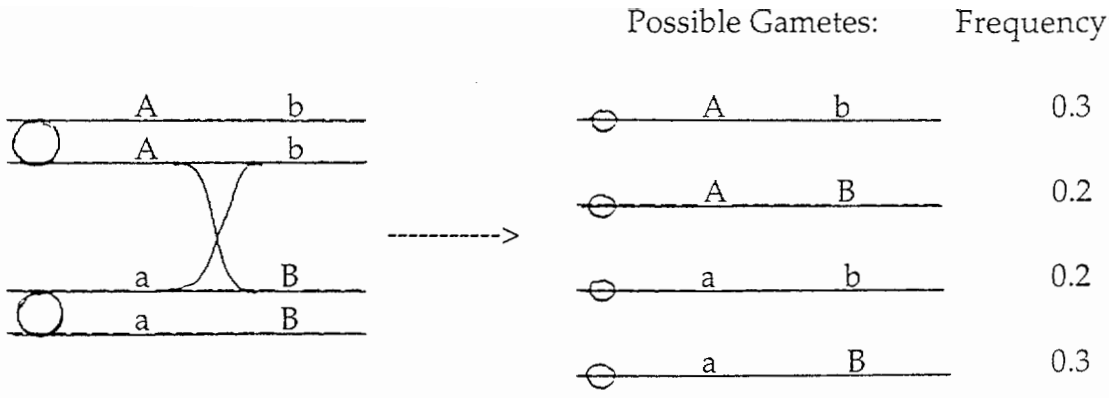
Calling the two linked genes A and B:



You now inter-cross F_1 mice to produce F_2 progeny.

(b) What fraction of the F_2 mice that are born will look like wild-type? Remember you must take into account both recombination and the lethality of double mutants. This problem should make it clear why mapping studies are, whenever possible, done by crossing heterozygous individuals to a homozygous recessive test strain.

Because the two genes are linked, they will not assort independently in the F_2. Therefore, the relative frequencies of gametes produced by the heterozygous F_1 parents must be determined. Crossover classes, during and after meiosis I, are drawn on the next page (applies for both parents). Since the two genes are 40 cM apart, each crossover class should appear in 20% of total gametes.



A Punnett square can then be used to represent all possible combinations of gametes, and their relative frequencies, in the F_2.

	AB 0.2	Ab 0.3	aB 0.3	ab 0.2
AB 0.2	**AABB** **0.04**	**AABb** **0.06**	**AaBb** **0.06**	**AaBb** **0.04**
Ab 0.3	**AAbB** **0.06**	AAbb 0.09	**AAbB** **0.09**	Aabb 0.06
aB 0.3	**aABB** **0.06**	**aABb** **0.09**	aaBB 0.09	aaBb 0.06
ab 0.2	**aAbB** **0.04**	aAbb 0.06	aabB 0.06	aabb 0.04

Wild-type mice are in bold. F_2 progeny with genotype aabb are dead and do not figure into the percentage of mice that are phenotypically normal.

$$\text{fraction wild-type} = \frac{(4 + 6 + 6 + 4 + 6 + 9 + 6 + 9 + 4)}{96} = \frac{54}{96} = 56.25\%$$

2. Consider the three hypothetical recessive traits in Drosophilia: bent wings (bt), yellow body (yl), and long bristles (lg). You are given a fly that looks like wild-type, but is heterozygous for all three traits. You set up a test cross of your to a that is homozygous recessive for all three traits. From 500 progeny the following types are found (abbreviations are used for recessive traits and the corresponding dominant traits are indicated by a +).

Phenotype	number
bt yl lg	34
+ + +	46
bt + +	13
+ yl lg	7
bt yl +	202
+ + lg	193
bt + lg	3
+ yl +	2

(a) If the fly that you were given came from a cross of two true-breeding lines, what would be the phenotypes of the two true-breeding lines?

The two true breeding lines would have the phenotypes of the 2 parental classes. The two parental classes are the most common. Therefore the two true breeding lines are

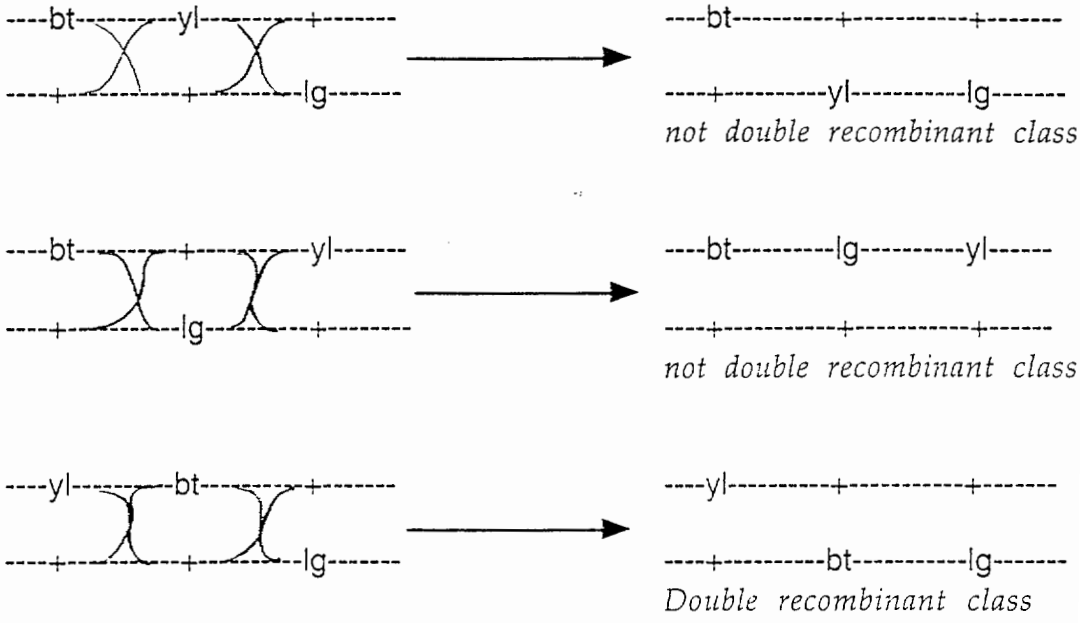
bt yl + / bt yl + *and* + + lg / + + lg

(b) Produce a genetic map showing the order and the relative positions of the genes that control the three traits.

To determine the gene order it is necessary to determine the double recombinant classes. The double recombinant classes are the lest frequent classes. They correspond to

bt + lg *and* + yl +

The order that will give you the double recombinant classes when two recombination events occur is the correct order.



The order is yl--bt--lg

The distance between yl and bt:
recombinant classes for yl and bt

bt + +	and	bt + lg
+ yl lg		+ yl +

distance = rec/total x 100 = (13+7+3+2)/500 x 100 = 5 cM

The distance between bt and lg:
recombinant classes for bt and lg

bt yl lg	and	bt + lg
+ + +		+ yl +

distance = rec/total x 100 = (34+46+3+2)/500 x 100 = 17 cM

```
----yl--------bt----------------lg-------
    |--5cM--|--   17cM   --|
```

3. You have isolated three new yeast mutants that are copper-resistant . Mutant 1 is dominant for resistance and is in mating type α. Mutants 2 and 3 are recessive and are in mating type **a**. You mate mutant 1 to mutant 2, sporulate the resulting diploid and dissect tetrads. Out of 25 tetrads, 4 have two copper resistant spores and two copper sensitive (wild-type) spores, 5 have four copper-resistant spores, and 16 have three copper-resistant spores and one wild-type spore.

(a) What can you conclude about the relative positions of mutation 1 and mutation 2 ?

mut1R,mut2S x mut1S,mut2R

----mu1R-----mut2S----
----mu1S----- mut2R----

	PD	*T*	*NPD*
	$mut1^R,mut2^S$	$mut1^R,mut2^S$	$mut1^R,mut2^R$
	$mut1^R,mut2^S$	$mut1^S,mut2^S$	$mut1^R,mut2^R$
	$mut1^S,mut2^R$	$mut1^R,mut2^R$	$mut1^S,mut2^S$
	$mut1^S,mut2^R$	$mut1^S,mut2^R$	$mut1^S,mut2^S$
spore ratio	*4 rest : 0 sens*	*3 rest : 1 sens*	*2 rest : 2 sens*
# of tet.	*5*	*16*	*4*

ratio of PD : T : NPD is 1 : 4 : 1 suggesting the two genes are unlinked

(b) If instead of dissecting tetrads you had examined 100 random spores, what fraction of the spores would you expect to be copper-resistant? What fraction of the copper-resistant spores would you expect to produce copper-resistant diploids when mated to wild-type?

The genes are unlinked so they undergo independent assortment i.e. you have 4 equally likely genotypes:

$mut1^R,mut2^S$ $mut1^S,mut2^R$ $mut1^R,mut2^R$ $mut1^S,mut2^S$

The first three genotypes have a copper-resistant phenotype while the other is copper-sensitive, therefore, 3/4 will be copper resistant and 1/4 will be sensetive. Only the spores that are $mut1^R$ *will give rise to copper-resistant diploids. Only the 1st and 3rd genotypes are* $mut1^R$*, therefore, only 1/2 will give rise to resistant diploids.*

Next, you mate mutant 1 to mutant 3, sporulate the diploid and dissect tetrads. Of 25 tetrads, 21 have four copper-resistant spores, and 4 have three copper-resistant spores and one wild-type spore.

(c) What can you conclude about the relative positions of mutation 1 and mutation 3 ?

$mut1^R,mut3^S$ x $mut1^S,mut3^R$

----$mu1^R$----- $mut3^S$----
----$mu1^S$----- $mut3^R$----

	PD	*T*	*NPD*
	$mut1^R,mut3^S$	$mut1^R,mut3^S$	$mut1^R,mut3^R$
	$mut1^R,mut3^S$	$mut1^S,mut3^S$	$mut1^R,mut3^R$
	$mut1^S,mut3^R$	$mut1^R,mut3^R$	$mut1^S,mut3^S$
	$mut1^S,mut3^R$	$mut1^S,mut3^R$	$mut1^S,mut3^S$
spore ratio	*4 rest : 0 sens*	*3 rest : 1 sens*	*2 rest : 2 sens*
# of tet.	*21*	*4*	*0*

ratio of PD : T : NPD is not 1 : 4 : 1 suggesting the two genes are linked
distane = 100 x # tetratypes/(2 x total tetrads) = 4 / (2 x 25) = 8 cM

(d) If instead of dissecting tetrads you had examined 100 random spores, what fraction of the spores would you expect to be copper-resistant? What fraction of the copper-resistant spores would you expect to produce copper-resistant diploids when mated to wild-type?

The genes are inked so they do not undergo independent assortment i.e. you have 4 genotypes of varying genotypes:

$mut1^R,mut3^S$ $mut1^S,mut3^R$ $mut1^R,mut3^R$ $mut1^S,mut3^S$

The first two genotypes are parental classes and the other two are recombinant classes. When two genes are 8 cM apart then 92% of the offspring will be of the patental classes while 8% will be recombinant. Therefore the fraction of each genotype is

$mut1^R,mut3^S$	$mut1^S,mut3^R$	$mut1^R,mut3^R$	$mut1^S,mut3^S$
46%	46%	4%	4%

The first three genotypes have a copper-resistant phenotype while the other is copper-sensitive, therefore, 96% will be copper resistant and 4% will be sensetive. Only the spores that are $mut1^R$ will give rise to copper-resistant diploids. Only the 1st and 3rd genotypes are $mut1^R$, therefore, only 50%will give rise to resistant diploids.

7.03 Problem Set 3

due before 5 PM on Thursday, October 9

Hand in answers in recitation section or in the box outside the class

1. You have isolated two new clear-plaque mutants of phage λ designated **cI-1** and **cI-2**. The **mi**$^-$ mutant gives small plaques. You cross a **cI-1 mi**$^-$ double mutant to a **cI-2** mutant by coinfecting host bacteria with phage of both types. After one round of growth you plate the resulting phage and examine the plaque types. Out of 1000 plaques, 20 have turbid plaques while the rest of the plaques are clear. Of the 20 turbid plaques, 15 are small and 5 are of the normal large size. Produce a genetic map showing the relative order of **cI-1**, **cI-2**, and **mi**$^-$ mutations and the distances between mutations in map units..

2. You have isolated a frameshift mutation in the T4 **rII** gene. By treating the mutant with proflavine you are able to isolate 20 **rII**$^+$ revertants of your mutation. By backcrossing to wild-type T4 you are able to show that all 20 revertants contain a second **rII**$^-$ mutation. All of these mutations map close to the original frameshift but to your surprise you notice that the suppressors mutations are not distributed randomly around the position of the original frameshift mutation — they all map on one side of the original mutation. What is the probable cause for this bias in position of the suppressor mutations (there is actually more than one possible explanation).

3. The sequencing of whole genomes is a rapidly developing area in modern genetics. Large-scale sequencing projects have the potential to define new genes without the need for isolating mutations in these genes. However, identifying genes from sequence information has pitfalls and distinguishing genuine coding sequences from fortuitous open reading frames (ORFs) can be more difficult than it might at first seem. If we define ORFs of at least 100 codons to be of sufficient length to be significant, it is interesting to calculate how often an ORF of this length would appear in a random (noncoding) sequence of DNA.

(a) Calculate the probability of finding an ORF of 100 codons with no stop codons in random DNA. Now determine how many such chance ORFs would appear on average in a bacterial genome of 10 Mbp.

(b) Often bacteria have skewed ratios of G-C and A-T base pairs. For a bacteria whose DNA is 70% G-C calculate the probability of a sequence of 100 codons with no stop codons. (Hint: use the content of A-T base pairs in stop codons to recalculate the average probability of a stop). Again, determine how many such chance ORFs would appear on average in a bacterial genome of 10 Mbp.

(c) For a bacteria whose DNA is 70% A-T calculate the probability of finding a sequence of 100 codons with no stop codons. Determine how many such chance ORFs would appear on average in a bacterial genome of 10 Mbp.

7.03 Problem Set 3 Solutions!

due before 5 PM on Thursday, October 9

Hand in answers in recitation section or in the box outside the class

1. You have isolated two new clear-plaque mutants of phage λ designated **cI-1** and **cI-2**. The **mi⁻** mutant gives small plaques. You cross a **cI-1 mi⁻** double mutant to a **cI-2** mutant by coinfecting host bacteria with phage of both types. After one round of growth you plate the resulting phage and examine the plaque types. Out of 1000 plaques, 20 have turbid plaques while the rest of the plaques are clear. Of the 20 turbid plaques, 15 are small and 5 are of the normal large size. Produce a genetic map showing the relative order of **cI-1**, **cI-2**, and **mi⁻** mutations and the distances between mutations in map units.

The parental genotypes are **cI-1 cI+2 mi-** and **cI+1 cI-2 mi+**. The two possible orders are:

cI-1 cI+2 mi-

or

cI+2 cI-1 mi-

The rarest class gives turbid, normal sized plaques, so their genotype must be **cI+1 cI+2 mi+**. They are one of the double crossover recombinants. To get this genotype from a double crossover of the parental genotypes, the order must be **cI-1 cI+2 mi-**.

The number of recombinants resulting from a crossover between **cI-1** and **cI-2** is (15 + 5)*2. So the distance in map units is (40/1000) *100 = 4 m.u.

Out of 20 turbid plaques, only 5 are normal sized colonies, so they are the result of a crossover between **cI-2** and **mi**. The distance between **cI-2** and **mi** is determined by the same formula, m.u. = (recombinants/total)*100. In this case, there are 5 recombinants out of a total of 20 turbid plaques. So the distance in map units = (5/20)*100 = 25 m.u.

2. You have isolated a frameshift mutation in the T4 **rII** gene. By treating the mutant with proflavine you are able to isolate 20 **rII⁺** revertants of your mutation. By backcrossing to wild-type T4 you are able to show that all 20 revertants contain a second **rII⁻** mutation. All of these mutations map close to the original frameshift but to your surprise you notice that the suppressors mutations are not distributed randomly around the position of the original frameshift mutation — they all map on one side of the original mutation. What is the probable cause for this bias in position of the suppressor mutations (there is actually more than one possible explanation).

* Mutation 1 produces a stop codon immediately downstream. So downstream mutations will not restore the wildtype phenotype . The transcription is still truncated. So the phenotype will still be rII⁻.

* Any revertant frameshift upstream of the first mutation results in a stop codon, so the phenoytpe will still be rII⁻.

*The first mutation borders on an important domain. Any mutation in the domain, although it will change the frame, will destroy the protein function, so the phenotype will still be rII⁻.

*The first mutation is very close to either the start or the stop codon so mutations on one side do not change the frame in which the translation occurs.

3. The sequencing of whole genomes is a rapidly developing area in modern genetics. Large-scale sequencing projects have the potential to define new genes without the need for isolating mutations in these genes. However, identifying genes from sequence information has pitfalls and distinguishing genuine coding sequences from fortuitous open reading frames (ORFs) can be more difficult than it might at first seem. If we define ORFs of at least 100 codons to be of sufficient length to be significant, it is interesting to calculate how often an ORF of this length would appear in a random (noncoding) sequence of DNA.

(a) Calculate the probability of finding an ORF of 100 codons with no stop codons in random DNA. Now determine how many such chance ORFs would appear on average in a bacterial genome of 10 Mbp.

3 out of 64 codons are stop codons.

p(non-stop codon)= 61/64

p(no stop codon in 100 codons)= $(61/64)^{100}$= $8.22*10^{-3}$

average #of ORFs = $8.22*10^{-3}$ * 10 Mbp * 2
[multiply by a fator of 2 to account for both directions]
= $16.44*10^{4}$ ORFs.

(b) Often bacteria have skewed ratios of G-C and A-T base pairs. For a bacteria whose DNA is 70% G-C calculate the probability of a sequence of 100 codons with no stop codons. (Hint: use the content of A-T base pairs in stop codons to recalculate the average probability of a stop). Again, determine how many such chance ORFs would appear on average in a bacterial genome of 10 Mbp.

35% G, 35% C, 15% A, 15% T
The stop codons are UAA, UAG, AGA.
p(stop codon)= (0.15*0.15*0.15 + 0.15*0.15*0.35 + 0.15*0.35*0.15)
= 0.019125

p(non-stop codon)= 1-0.019125

p(no stop codon for 100 codons)= $(1-0.019125)^{100}$ = 0.145

average # of ORFs = 0.145 * 10 Mbp * 2 = $2.90*10^{6}$ ORFs

(c) For a bacteria whose DNA is 70% A-T calculate the probability of finding a sequence of 100 codons with no stop codons. Determine how many such chance ORFs would appear on average in a bacterial genome of 10 Mbp.

35% A, 35% T, 15% C, 15% G.
p(stop codon)= (0.35*0.35*0.35 + 0.35*0.35*0.15 + 0.35*0.15*0.35)
= 0.079625

p(non-stop codon)= 1- 0.79625

p(no stop codons for 100 codons)= $(1- 0.79625)^{100} = 2.49*10^{-4}$

average # of ORFs= $2.49*10^{-4}$ * 10 Mbp * 2 = $4.98*10^{3}$ ORFs

7.03 Problem Set 4

due before 5 PM on Tuesday, October 21

Hand in answers in recitation section or in the box outside the class

1. Phage P1 is grown on an *E. coli* host that is pyrF$^-$. The phage produced are then used to infect a hisB$^-$ argE$^-$ recipient strain. His$^+$ transductants are selected and then tested for the presence of the other markers. The phenotypes of 500 His$^+$ transductants are given below.

Phenotype	Number
Pyr$^+$ Arg$^+$	15
Pyr$^+$ Arg$^-$	150
Pyr$^-$ Arg$^+$	235
Pyr$^-$ Arg$^-$	100

(a) Are these three genes within 10^5 base pairs (the size of phage P1 genome) of one another? Justify your answer.

(b) Draw a map giving the relative order and the cotransductional distances between the markers pyrF, hisB, and argE.

(c) If Arg$^+$ transductants had been selected (instead of His$^+$), what fraction of these would be His$^+$? What fraction would be Pyr$^+$?

2. **(a)** Consider a portion of the *E. coli* chromosome that contains three genetic markers in the following order: metD — proA — lacZ. Starting with an F$^+$, metD$^+$, proA$^+$, lacZ$^+$, Str^s strain, you isolate an Hfr that transfers metD$^+$ early and efficiently but transfers the proA$^+$ and LacZ$^+$ markers very inefficiently. Draw a map of the relevant portion of the *E. coli* chromosome in the Hfr showing the integrated F (indicate both the site of integration and the orientation of the origin of transfer).

(b) If the P1 cotransductional distance between metD and proA is 70% in a wild-type *E. coli* strain, what will the cotransduction frequency between metD and proA be in the Hfr strain isolated in part (**a**).

(c) You isolate an F' from the Hfr described in part (**a**) by mating to an F$^-$, metD$^-$, proA$^-$, lacZ$^-$, Str^r strain and selecting for lacZ$^+$ and Str^r after a brief period of mating. This F' strain is proA$^+$ and metD$^-$. Describe the positions of the sequences that recombined to produce the F'.

3. **(a)** Phenyl β,D-galactoside (PG) is a substrate for β-Galactosidase, but is not an inducer of the *lac* operon. Isopropyl β,D-thiogalactoside (IPTG) is an inducer of the *lac* operon, but not a substrate for β-Galactosidase. Can wild-type cells grow on media containing PG? IPTG? PG+IPTG? Explain.

Following is a positional map of the *lac* operon,

I P O Z Y

where *P* represents the promoter region necessary for transcription of the *Z* (β-Galactosidase), and *Y* (Permease) cistrons. One can isolate mutations in this promoter region (P^-), which prevent the promoter from interacting with the RNA polymerase holoenzyme, thus blocking transcription.

(b) Given what you know about the P^- mutation and other mutations within the *lac* operon. give the induction state of the following strains under inducing and non-inducing conditions. Justify your answers.

		LacZ induction	
		GLUCOSE	IPTG
EXAMPLE	$I^+ P^+ O^+ Z^+ Y^+$	-	+
	$I^+ P^- O^+ Z^+ Y^+ / F' I^+ P^+ O^+ Z^+ Y^+$		
	$I^+ P^- O^+ Z^+ Y^+ / F' I^+ P^+ O^+ Z^- Y^-$		
	$I^+ P^- O^c Z^+ Y^+ / F' I^+ P^+ O^+ Z^+ Y^+$		
	$I^+ P^- O^+ Z^+ Y^+ / F' I^+ P^+ O^c Z^+ Y^+$		

(c) You isolate an unidentified mutant (X^-), that fails to induce *LacZ* expression on IPTG plates. You generate the following strains, and assay them for induction of *LacZ* under inducing and noninducing conditions.

	LacZ induction	
	GLUCOSE	IPTG
X^-	-	-
X^-/ F' I^+ P^+ O^+ Z^+ Y^+	-	+
X^-/ F' I^+ P^- O^+ Z^+ Y^+	-	-
X^-/F' I^{-d} P^+ O^+ Z^+ Y^+	+	+
X^-/F' I^+ P^+ O^c Z^- Y^+	-	+
X^-/F' I^+ P^- O^+ Z^+ Y^+	-	-

(This last strain is repeated information)

Where in the *lac* operon does the mutation X^- likely reside? Explain how each of the strains above support your conclusion.

7.03 Problem Set 4
Solutions

due before 5 PM on Tuesday, October 21
Hand in answers in recitation section or in the box outside the class

1. Phage P1 is grown on an *E. coli* host that is pyrF$^-$. The phage produced are then used to infect a hisB$^-$ argE$^-$ recipient strain. His$^+$ transductants are selected and then tested for the presence of the other markers. The phenotypes of 500 His$^+$ transductants are given below.

Phenotype	Number
Pyr$^+$ Arg$^+$	15
Pyr$^+$ Arg$^-$	150
Pyr$^-$ Arg$^+$	235
Pyr$^-$ Arg$^-$	100

(a) Are these three genes within 10^5 base pairs (the size of phage P1 genome) of one another? Justify your answer.

Yes, the three genes are within 10^5 base pairs of one another, since one can see cotransduction of all three markers. The upper limit for cotransduction is 10^5 base pairs, which is the amount of DNA that can be stored in phage P1.

(b) Draw a map giving the relative order and the cotransductional distances between the markers pyrF, hisB, and argE.

If we look at the least frequent class (His$^+$ Pyr$^+$ Arg$^+$), we can see that this can be genrated by a quadruple recombination if the orientation of the genes is:

hisB--pyrF--argE

The distances between hisB and the other markers are determined by the frequency of marker cotransduction. Thus the distance between hisB and argE is:

***(15+235)/500=50%** cotransduction frequency.*

The distance between hisB and pyrF is:

***(235+100)/500=67%** cotransduction frequency.*

*Given the data available in this problem, there is no way to generate a completely accurate distance (as a cotransduction frequency) between argE and pyrF. The only accurate statement is that the distance between argE and pyrF is **> 50%** cotransduction frequency*

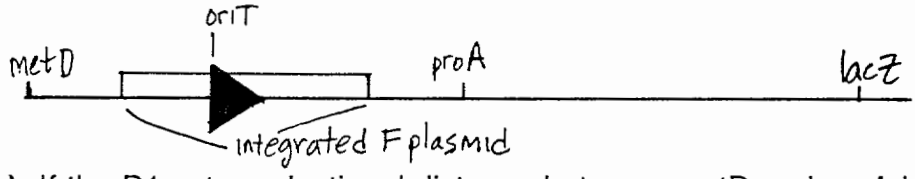
*There are a number of ways to generate an **estimated** value for the distance between argE and pyrF. If one considers cotransduction between Pyr$^-$ and Arg$^+$, one can generate a frequency of*

cotransduction of ***235/(235+100)=78%. This value is not completely accurate, because it fails to take into account the events which cotransduce pyrF⁻ and argE⁺, but do not include the hisB⁺ allele.***

(c) If Arg⁺ transductants had been selected (instead of His⁺), what fraction of these would be His⁺? What fraction would be Pyr⁺?

This problem relies on the cotransductional distances calculated in part (c). If the cotransductional freqency between hisB and argE is 50%, then the frequency of Arg⁺ transductants that are also His⁺ will be ***50%.***
The accurate answer for the frequency of Arg⁺ transductants that are also Pyr⁺ is ***<50%*** *(since the pyrF⁻ allele is being cotransduced with the argE⁺ allele at a frequency >50%).*
Using our estimated value for the cotransductional distance between argE and pyrF, (78%) we can generate an estimated value for a frequency of Arg + transductants that are also Pyr+ = 22%

2. **(a)** Consider a portion of the *E. coli* chromosome that contains three genetic markers in the following order: metD — proA — lacZ. Starting with an F⁺, metD⁺, proA⁺, lacZ⁺, Str^s strain, you isolate an Hfr that transfers metD⁺ early and efficiently but transfers the proA⁺ and LacZ⁺ markers very inefficiently. Draw a map of the relevant portion of the *E. coli* chromosome in the Hfr showing the integrated F (indicate both the site of integration and the orientation of the origin of transfer).

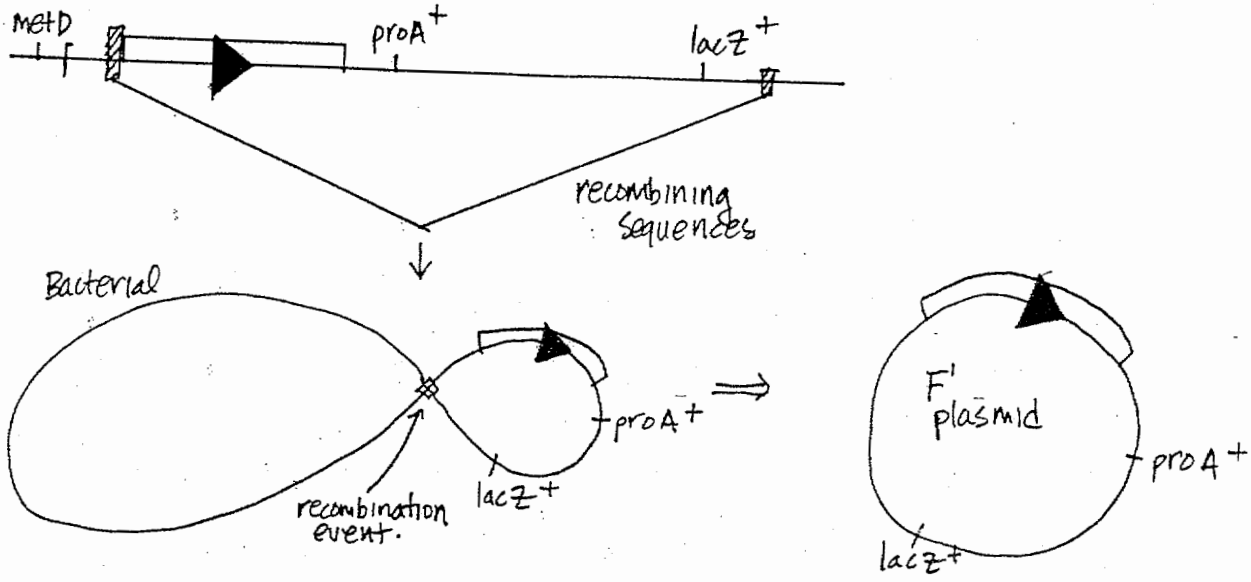


(b) If the P1 cotransductional distance between metD and proA is 70% in a wild-type *E. coli* strain, what will the cotransduction frequency between metD and proA be in the Hfr strain isolated in part (**a**).

The size of the F plasmid inserted between metD and proA is 10^5 base pairs. The cotransduction frequency between metD and proA would go down to 0% in the isolated Hfr since they would be seperated by more than 10^5 base pairs.

(c) You isolate an F' from the Hfr described in part (a) by mating to an F^-, $metD^-$, $proA^-$, $lacZ^-$, Str^r strain and selecting for $lacZ^+$ and Str^r after a brief period of mating. This F' strain is $proA^+$ and $metD^-$. Describe the positions of the sequences that recombined to produce the F'.

The recombination event that produces the isolated F' is as diagrammed below:



3. **(a)** Phenyl β,D-galactoside (PG) is a substrate for β-Galactosidase, but is not an inducer of the *lac* operon. Isopropyl β,D-thiogalactoside (IPTG) is an inducer of the *lac* operon, but not a substrate for β-Galactosidase. Can wild-type cells grow on media containing PG? IPTG? PG+IPTG? Explain.

The cells cannot grow on media containing PG alone, since they will not be able to activate the lac operon. On IPTG, the cells will not grow because they cannot use IPTG as a substrate for β-galactosidase. On media containing both PG and IPTG, the cells will be able to grow, since IPTG can activate the lac operon, and PG can act as a subtrate for β-galactosidase.

Following is a positional map of the *lac* operon,

I P O Z Y

where *P* represents the promoter region necessary for transcription of the *Z* (β-Galactosidase), and *Y* (Permease) cistrons. One can isolate mutations in this promoter

region (P^-), which prevent the promoter from interacting with the RNA polymerase holoenzyme, thus blocking transcription.

(b) Given what you know about the P^- mutation and other mutations within the *lac* operon. give the induction state of the following strains under inducing and non-inducing conditions. Justify your answers.

		LacZ induction	
		GLUCOSE	IPTG
EXAMPLE	$I^+ P^+ O^+ Z^+ Y^+$	-	+
	$I^+ P^- O^+ Z^+ Y^+ / F' I^+ P^+ O^+ Z^+ Y^+$	-	+
	$I^+ P^- O^+ Z^+ Y^+ / F' I^+ P^+ O^+ Z^- Y^-$	-	-
	$I^+ P^- O^c Z^+ Y^+ / F' I^+ P^+ O^+ Z^+ Y^+$	-	+
	$I^+ P^- O^+ Z^+ Y^+ / F' I^+ P^+ O^c Z^+ Y^+$	+	+

(c) You isolate an unidentified mutant (X^-), that fails to induce *LacZ* expression on IPTG plates. You generate the following strains, and assay them for induction of *LacZ* under inducing and noninducing conditions.

	LacZ induction	
	GLUCOSE	IPTG
X^-	-	-
$X^- / F' I^+ P^+ O^+ Z^+ Y^+$	-	+
$X^- / F' I^+ P^- O^+ Z^+ Y^+$	-	-
$X^- / F' I^{-d} P^+ O^+ Z^+ Y^+$	+	+
$X^- / F' I^+ P^+ O^c Z^- Y^+$	-	+

Where in the *lac* operon does the mutation X^- likely reside? Explain how each of the strains above support your conclusion.

The X mutation leads to an uninducible phenotype. Thus X is cannot be I^d, O^c or I^-. The second strain, $X^- / F' I^+ P^+ O^+ Z^+ Y^+$, shows that the X mutation is recessive. X cannot be I^s, which is dominant. Finally, the $X^- / F' I^+ P^+ O^c Z^- Y^+$ strain shows that X complements a Z mutation, thus X cannot be either a P^- or a Z^- mutation. Therefore, by elimination, the X mutation is likely to be a Y mutation in the permease gene. The other strains are consistent with the hypothesis that X represents a mutation in the permease gene.

PROBLEM SET 5

Due by 5 PM on October 30

1. Toxicologists digging at a chemical waste dump have identified a novel bacterium that can detoxify mercury-containing compounds. Biochemical experiments indicate that the bacterium produces an enzyme called curase only when mercury is present. You work in a genetics lab that has isolated mutants in three separate complementation groups, called merA, merB, and merC. Listed below are the curase levels of these strains in the presence and absence of mercury.

Strain	Curase Level (-) mercury	(+) mercury
1. merA+merB+merC+	0	200
2. merA-merB+merC+	200	200
3. merA+merB-merC+	200	200
4. merA+merB+merC-	0	0

a.) Which of these genes encodes for curase? What type of regulation appears to be used for the enzyme?

Since you can determine nothing further from the single mutant strains, you make merodiploids to perform cis-trans tests.

Strain	Curase Level (-) mercury	(+) mercury
5. F' A-B-C+ / F- A+B+C+	200	400
6. F' A-B+C+ / F- A+B-C+	200	400
7. F' A+B-C- / F- A+B+C+	0	200
8. F' A+B-C+ / F- A+B+C-	200	200
9. F' A-B+C- / F- A+B+C+	0	200
10. F' A-B+C+ / F- A+B+C-	0	200

b.) Which gene is a regulatory factor that acts in trans with respect to merC? Which gene is a regulatory sequence that acts in cis with merC? For both answers, give the numbers of the strains that allowed you to make the distinction.

c.) Based on the new data, propose a model for how merA, merB, and merC interact to regulate curase expression.

A new mutant is isolated that shows uninducible curase activity even when mercury is present. The mutation maps to a gene, called merD, that is linked to but genetically distinct from A, B, and C. To further determine the role of merD, you make several double mutant strains and test for curase levels.

Strain	Curase Level (-) mercury	(+) mercury
11. A+B+C+D-	0	0
12. A+B-C+D-	200	200
13. A-B+C+D-	200	200
14. A-B-C+D+	200	200

d.) What do these strains indicate about epistatic relationships between merA, B, and D? Draw a genetic regulatory pathway consistent with these results.

e.) You learn that merD is a trans-acting factor. Based on this new information, predict the expected levels of curase for these merodiploids. (Hint: look carefully at the curase levels in part b)

Strain	Curase Level (-) mercury	(+) mercury
14. F' A-B+C+D+ / F- A+B+C+D-	?	?
15. F' A+B-C-D+ / F- A+B+C+D-	?	?
16. F' A+B-C+D- / F- A-B+C+D-	?	?

f.) A dominant, gain-of-function mutation in merD is discovered and named merD*. Strains with this mutation show constitutive expression of curase. Explain this result by presenting a molecular model of how merD* affects curase expression.

Note: Question #2 was purposely omitted.

3. E.coli is normally unable to live solely on ramen noodles. However, you note that upon the addition of a colorless compound A, the cells, which are normally white, eventually turn blue and are able to live on ramen. Cloning by function has resulted in identifying a gene, called ramenase, which allows cells to grow on ramen noodles. Interested in the regulation of ramenase expression, you proceed to isolate several mutants.

First, you isolate mutants which do not grow on ramen even when compound A is added: mutants X, Y, Z.

	+A	
WT	blue, +	('+'indicates growth on ramen, '-' indicates no growth)
X^-	red, -	
Y^-	yellow, -	
Z^-	white, -	
X^-Y^-	yellow, -	
X^-Z^-	white, -	
Y^-Z^-	white, -	

a) Which gene can possibly be a permease?

b) Based on the above data, diagram the biochemical pathway for the metabolism of compound A (include colors). Indicate any ambiguities if there are insufficient data.

Oddly, some mutants which do not grow even in the presence of compound A still turn blue: mutants T and W. Subsequent analysis reveals that mutant W has premature stop codon in the ramenase gene itself.

	-A	+A
T^-	white, -	blue, -
W^-	white, -	blue, -
$\frac{T^-W^+}{T^+W^-}$	white, -	blue, +
$\frac{T^-W^-}{T^+W^+}$	white, -	blue, +

c) Diagram the complete regulation of ramenase through compound A. Account for all the data given.

Lastly, you screen for mutants that can live on ramen without the addition of compound A, and find only one: mutant V. Mapping experiments indicate that the mutation in this mutant maps extremely close to the mutation in T.

	-A	+A
V^-	white, +	blue, +
V^-W^-	white, -	white, -
V^-X^-	white, +	red, +
V^-Y^-	white, +	yellow, +
V^-Z^-	white, +	white, +
$\frac{V^-W^-}{V^+W^+}$	white, +	blue, +
$\frac{V^-W^+}{V^+W^-}$	white, +	blue, +

d) What is the most likely explanation for the behavior of mutant V?

PROBLEM SET 5 **ANSWERS**

1. Toxicologists digging at a chemical waste dump have identified a novel bacterium that can detoxify mercury-containing compounds. Biochemical experiments indicate that the bacterium produces an enzyme called curase only when mercury is present. You work in a genetics lab that has isolated mutants in three separate complementation groups, called merA, merB, and merC. Listed below are the curase levels of these strains in the presence and absence of mercury.

Strain	Curase Level	
	(-) mercury	(+) mercury
1. merA+merB+merC+	0	200
2. merA-merB+merC+	200	200
3. merA+merB-merC+	200	200
4. merA+merB+merC-	0	0

a.) Which of these genes encodes for curase? What type of regulation appears to be used for the enzyme?

The merC- mutant shows uninducible expression of curase. Therefore, ***merC*** *encodes for* ***curase.***
The merA- and merB- mutants shows constitutive expression of curase. This suggests ***negative regulation*** *of curase transcription.*

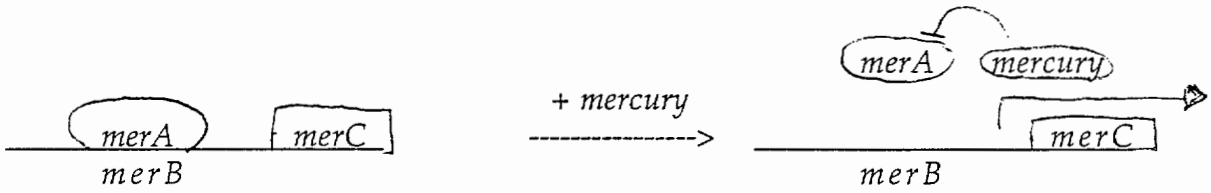
Since you can determine nothing further from the single mutant strains, you make merodiploids to perform cis-trans tests.

Strain	Curase Level	
	(-) mercury	(+) mercury
5. F' A-B-C+ / F- A+B+C+	200	400
6. F' A-B+C+ / F- A+B-C+	200	400
7. F' A+B-C- / F- A+B+C+	0	200
8. F' A+B-C+ / F- A+B+C-	200	200
9. F' A-B+C- / F- A+B+C+	0	200
10. F' A-B+C+ / F- A+B+C-	0	200

b.) Which gene is a regulatory factor that acts in trans with respect to merC? Which gene is a regulatory sequence that acts in cis with merC? For both answers, give the numbers of the strains that allowed you to make the distinction.

Strains 7 and 8 show that ***merB*** *is a* ***cis-acting*** *sequence.*
Strains 9 and 10 show that ***mer A*** *is a* ***trans-acting*** *regulatory factor.*
Since curase expression is negatively regulated, merA could be a repressor that binds to the merB operator site.

c.) Based on the new data, propose a model for how merA, merB, and merC interact to regulate curase expression.



OFF *ON*

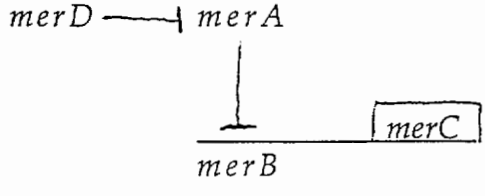
A new mutant is isolated that shows uninducible curase activity even when mercury is present. The mutation maps to a gene, called merD, that is linked to but genetically distinct from A, B, and C. To further determine the role of merD, you make several double mutant strains and test for curase levels.

Strain	Curase Level (-) mercury	(+) mercury
11. A+B+C+D-	0	0
12. A+B-C+D-	200	200
13. A-B+C+D-	200	200
14. A-B-C+D+	200	200

d.) What do these strains indicate about epistatic relationships between merA, B, and D? Draw a genetic regulatory pathway consistent with these results.

Strain 12 shows that ***merB is epistatic to merD****, since the double mutant has the phenotype of the merB single mutant. By the same logic, strain 13 shows that* ***merA is epistatic to merD.*** *Nothing can be determined from these four strains about the epistatic relationship of mer A and mer B because both of these single mutants have the same phenotype. However, we know from previous information that A is a repressor and B is an operator.*

Genetic regulatory pathway:



e.) You learn that merD is a trans-acting factor. Based on this new information, predict the expected levels of curase for these merodiploids. (Hint: look carefully at the curase levels in part b)

Strain	Curase Level (-) mercury	Curase Level (+) mercury
14. F' A-B+C+D+ / F- A+B+C+D-	0	400
The C gene is regulated as in wild-type. Since there are two copies, however, expression in the presence of mercury will be doubled.		
15. F' A+B-C-D+ / F- A+B+C+D-	0	200
Only one functional copy of C is present, and it is regulated normally. Therefore, curase levels will be the same as in a wild-type haploid.		
16. F' A+B-C+D- / F- A-B+C+D-	200	200
The copy of C on the F' is constitutively expressed. The copy of C on the F- chromosome will always be repressed, since no functional D is present in the cell. Therefore, curase levels will always be 200 units.		

f.) A dominant, gain-of-function mutation in merD is discovered and named merD*. Strains with this mutation show constitutive expression of curase. Explain this result by presenting a molecular model of how merD* affects curase expression.

One possible explanation is as follows:

MerD is a repressor of merA. Normally, merD is inactive in the absence of mercury. When it binds mercury, it changes conformation and is able to bind to merA and prevent it from binding to DNA. MerD has a mutation such that it is always in the active conformation and can bind to the repressor even in the absence of mercury. Therefore, merC transcription is constitutive.*

3. E.coli is normally unable to live solely on ramen noodles. However, you note that upon the addition of a colorless compound A, the cells, which are normally white, eventually turn blue and are able to live on ramen. Cloning by function has resulted in identifying a gene, called ramenase, which allows cells to grow on ramen noodles. Interested in the regulation of ramenase expression, you proceed to isolate several mutants.

First, you isolate mutants which do not grow on ramen even when compound A is added: mutants X, Y, Z.

	+A	
WT	blue, +	('+'indicates growth on ramen, '-' indicates no growth)
X^-	red, -	
Y^-	yellow, -	
Z^-	white, -	
X^-Y^-	yellow, -	
X^-Z^-	white, -	
Y^-Z^-	white, -	

a) Which gene can possibly be a permease?

Gene Z. It is epistatic to all other genes. If compound A cannot enter the cell then no color changes will be observed and the cell remains white.

b) Based on the above data, diagram the biochemical pathway for the metabolism of compound A (include colors). Indicate any ambiguities if there are insufficient data.

Based on the data above, it is not clear how the genes X, Y, Z are regulated; one can only infer that they are present at specific steps of the metabolism. Hence, any model which includes regulation of genes X, Y, and Z presumes extraneous data not available.

Model 1:

A (colorless) --Z--> *yellow* --Y--> *red* --X--> *blue*

Upon entry into the cell, A immediately turns yellow.

Model 2:

A (colorless) --Z--> --?--> *yellow* --Y--> *red* --X--> *blue*

Upon entry into the cell, A remains colorless, but another unidentified gene product changes A into a yellow intermediate.

Oddly, some mutants which do not grow even in the presence of compound A still turn blue: mutants T and W. Subsequent analysis reveals that mutant W has premature stop codon in the ramenase gene itself.

	-A	+A
T-	white, -	blue, -
W-	white, -	blue, -
$\frac{T^-W^+}{T^+W^-}$	white, -	blue, +
$\frac{T^-W^-}{T^+W^+}$	white, -	blue, +

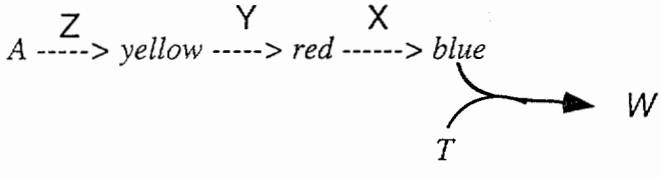
c) Diagram the complete regulation of ramenase through compound A. Account for all the data given.

There are two models which fit the data. From the above data, we can infer that T is an activator of W (ramenase). We can infer that the blue colored compound resulting fromthe metabolism of A is also an activator of W, because cells which are not blue cannot make W. However, we do not know the relationship between the blue colored compound and the T gene product.

Model 1:

$$A \xrightarrow{Z} yellow \xrightarrow{Y} red \xrightarrow{X} blue \dashrightarrow T \dashrightarrow W$$

Model 2:



Lastly, you screen for mutants that can live on ramen without the addition of compound A, and find only one: mutant V. Mapping experiments indicate that the mutation in this mutant maps extremely close to the mutation in T.

	-A	+A
V^-	white, +	blue, +
V^-W^-	white, -	blue, -
V^-X^-	white, +	red, +
V^-Y^-	white, +	yellow, +
V^-Z^-	white, +	white, +
$\frac{V^-W^-}{V^+W^+}$	white, +	blue, +
$\frac{V^-W^+}{V^+W^-}$	white, +	blue, +

d) What is the most likely explanation for the behavior of mutant V?

The mapping experiments mentioned in the text imply that mutant V and mutant T each have a mutation in the same gene. Since mutant V has constituitive expression of W, then it would seem that the mutant V results in the constituitive action of T gene product. Based on the two models of part c, mutant V would suggest that model 2 is more likely. (An analogous mutation would be $MalT^c$.)

7.03 Problem Set 4
Solutions

due before 5 PM on Tuesday, October 21
Hand in answers in recitation section or in the box outside the class

1. Phage P1 is grown on an *E. coli* host that is pyrF$^-$. The phage produced are then used to infect a hisB$^-$ argE$^-$ recipient strain. His$^+$ transductants are selected and then tested for the presence of the other markers. The phenotypes of 500 His$^+$ transductants are given below.

Phenotype	Number
Pyr$^+$ Arg$^+$	15
Pyr$^+$ Arg$^-$	150
Pyr$^-$ Arg$^+$	235
Pyr$^-$ Arg$^-$	100

(a) Are these three genes within 10^5 base pairs (the size of phage P1 genome) of one another? Justify your answer.

Yes, the three genes are within 10^5 base pairs of one another, since one can see cotransduction of all three markers. The upper limit for cotransduction is 10^5 base pairs, which is the amount of DNA that can be stored in phage P1.

(b) Draw a map giving the relative order and the cotransductional distances between the markers pyrF, hisB, and argE.

If we look at the least frequent class (His$^+$ Pyr$^+$ Arg$^+$), we can see that this can be genrated by a quadruple recombination if the orientation of the genes is:

hisB--pyrF--argE

The distances between hisB and the other markers are determined by the frequency of marker cotransduction. Thus the distance between hisB and argE is:

***(15+235)/500=50%** cotransduction frequency.*

The distance between hisB and pyrF is:

***(235+100)/500=67%** cotransduction frequency.*

*Given the data available in this problem, there is no way to generate a completely accurate distance (as a cotransduction frequency) between argE and pyrF. The only accurate statement is that the distance between argE and pyrF is **> 50%** cotransduction frequency*

*There are a number of ways to generate an **estimated** value for the distance between argE and pyrF. If one considers cotransduction between Pyr$^-$ and Arg$^+$, one can generate a frequency of*

You want to go into greater detail with the mutants that you have so you make strains that have your mutation plus a GAL80-VP16 fusion gene. GAL80-VP16 encodes the Gal80 protein with the activation domain of the powerful viral activator VP16 fused to GAL80 in frame. You observe the following phenotypes.

	GAL1 Levels
Mutant	-Galactose
wt, GAL80-VP16	1000
A-, GAL80-VP16	1000
B-, GAL80-VP16	1

(b) What does the following data tell us about the nature of the A- and B- mutations?

You then search for mutations that lead to constitutive expression of the GAL1 gene and find two mutants (C-, D-) that map to different locations in the genome. You mate these mutants to a WT strain and get the folowing phenotypes.

	GAL1 Levels	
Mutant	+Galactose	-Galactose
wt	100	1
C-	100	100
D-	100	100
C-/wt	100	100
D-/wt	100	1

(c) Describe the nature of the C- and D- mutants and where you think the mutations are.

(d) Fill in the levels of GAL1 activity you would expect from the folowing strains

Strain	+Galactose	-Galactose
C-, GAL80-VP16		
B-, C+ / B+, C-		
B-, D+ / B+, D-		
B-, D-		

3. A recessive blood disorder appears at a frequency of 2×10^{-5} within the population of a large country. The gene for this trait is known, and the allele that gives rise to the disorder can be detected by a DNA-based test. By screening a large number of individuals in the population, it is found that the allele frequency for the disorder is 2×10^{-3}.

(a) How do we know that this population is not in Hardy-Weinberg equilibrium for the disorder?

(b) What is the frequency of carriers in the population?

(c) A population that is not in Hardy-Weinberg equilibrium usually indicates that mating within the population is not occurring at random. Consider the possibility that all of the alleles for the disorder occur within a subgroup of the country's population, and that although there is random mating within the subgroup there is essentially no inter-mating between this subgroup and the larger population. Calculate the fraction of the total population that would need to be in the isolated subgroup in order to explain the measured frequency of the disorder and the allele frequency within the total population.

(d) In class we considered how a balance between selection and new mutations can set the allele frequency for autosomal alleles. We did not carry out an equivalent analysis of X-linked traits partly because the need to treat males and females separately, makes the calculations considerably more difficult. Nevertheless, there is a simple and important feature of the population genetics of X-linked traits, first derived by the famous statistician J. B. S. Haldane, that we will work out here. First, determine the fraction of all of the X chromosomes in a large population that are in males. (It may be helpful to check your answer by considering a population that includes say 500 males and 500 females).

(e) Consider an X-linked recessive disorder that has fitness = 0 in males and fitness = 1 in heterozygous females. (For example, prior to the improvements in medicine developed in this century, hemophilia A alleles would have had essentially these fitness properties). Given your answer for part **d**, what fraction of the alleles for the disorder will be lost each generation due to selection against affected males?

(f) In order for the allele frequency for the disorder to remain at steady-state, new alleles must arise by mutation at the same rate as alleles are lost by selection. Given this constraint, what fraction of the alleles for the disorder would you expect to be new alleles and what fraction would you expect to be inherited from carrier females?

7.03 Problem Set 6 Solutions

due before 5 PM on Thursday, November 13

Hand in answers in recitation section or in the box outside the class

2. You want to study regulation of the GAL1 gene. Like the good geneticist that you are, you screen for mutants that are defective in GAL1 transcription and isolate the following two mutants (A- & B-).

	GAL1 Levels	
Mutant	+Galactose	-Galactose
wt	100	1
A-	1	1
B-	1	1

To further study these mutants, you mate each mutant to a wt strain and to a gal4- strain.

	GAL1 Levels	
Mutant	+Galactose	-Galactose
gal4-	1	1
A-/gal4-	1	1
B-/gal4-	1	1
A-/wt	100	1
B-/wt	100	1

(a) Describe the A- and B- mutations and where you think they are.

A- and B- mutations are recessive mutations which gives an uninducible phenotype. Since they do not complement a gal4- mutation, they are probably mutations in the GAL4 gene.

You want to go into greater detail with the mutants that you have so you make strains that have your mutation plus a GAL80-VP16 fusion gene. GAL80-VP16 encodes the Gal80 protein with the activation domain of the powerful viral activator VP16 fused to GAL80 in frame. You observe the following phenotypes.

	GAL1 Levels
Mutant	-Galactose
wt, GAL80-VP16	1000
A-, GAL80-VP16	1000
B-, GAL80-VP16	1

(b) What does the following data tell us about the nature of the A- and B- mutations?

B- is either a complete deletion of the GAL4 gene or it is a mutation in the DNA binding domain. A- has to be a loss of function mutation in the activation domain of GAL4. The A- mutation, however, must still be capable to binding GAL80 and the UAS.

You then search for mutations that lead to constitutive expression of the GAL1 gene and find two mutants (C-, D-) that map to different locations in the genome. You mate these mutants to a WT strain and get the folowing phenotypes.

	GAL1 Levels	
Mutant	**+Galactose**	**-Galactose**
wt	100	1
C-	100	100
D-	100	100
C-/wt	100	100
D-/wt	100	1

(c) Describe the nature of the C- and D- mutants and where you think the mutations are.

C- is a dominant mutation in the activation domain of GAL4 such that GAL80 is no longer able to bind GAL4. D- is most likely a recessive loss of function in the GAL80 gene.

(d) Fill in the levels of GAL1 activity you would expect from the folowing strains

Strain	**+Galactose**	**-Galactose**
C-, GAL80-VP16	*100*	*100*
B-, C+ / B+, C-	*100*	*100*
B-, D+ / B+, D-	*100*	*1*
B-, D-	*1*	*1*

3. A recessive blood disorder appears at a frequency of 2×10^{-5} within the population of a large country. The gene for this trait is known, and the allele that gives rise to the disorder can be detected by a DNA-based test. By screening a large number of individuals in the population, it is found that the allele frequency for the disorder is 2×10^{-3}.

(a) How do we know that this population is not in Hardy-Weinberg equilibrium for the disorder?

If the alelle is in HW equilibrium then

$$q^2 = f(aa)$$

but, $q = 2 \times 10^{-3}$ *and* $f(aa) = 2 \times 10^{-5}$

$$(2 \times 10^{-3})^2 = 4 \times 10^{-6} \neq 2 \times 10^{-5}$$

(b) What is the frequency of carriers in the population?

$$f(Aa) = 2pq \sim 2q = 4 \times 10^{-3}$$

(c) A population that is not in Hardy-Weinberg equilibrium usually indicates that mating within the population is not occurring at random. Consider the possibility that <u>all</u> of the alleles for the disorder occur within a subgroup of the country's population, and that although there is random mating within the subgroup there is essentially no inter-mating between this subgroup and the larger population. Calculate the fraction of the total population that would need to be in the isolated subgroup in order to explain the measured frequency of the disorder and the allele frequency within the total population.

n = fraction of subgroup in population
q/n = freq of allele in the subgroup
f(aa)/n = freq. of disease in the subgroup
if the disease is in HW equilibrium in the subgroup then

$$(q/n)^2 = f(aa)/n$$

$$(4 \times 10^{-6})/n^2 = (2 \times 10^{-5})/n$$

$$n = 0.2$$

Subgroup is 20% of population.

(d) In class we considered how a balance between selection and new mutations can set the allele frequency for autosomal alleles. We did not carry out an equivalent analysis of X-linked traits partly because the need to treat males and females separately, makes the calculations considerably more difficult. Nevertheless, there is a simple and important feature of the population genetics of X-linked traits, first derived by the famous statistician J. B. S. Haldane, that we will work out here. First, determine the fraction of all of the X chromosomes in a large population that are in males. (It may be helpful to check your answer by considering a population that includes say 500 males and 500 females).

Females have 2 X chromosomes while males have 1 X chromosome therefore 1/3 of X chromosomes are in the male population.

(e) Consider an X-linked recessive disorder that has fitness = 0 in males and fitness = 1 in heterozygous females. (For example, prior to the improvements in medicine developed in this century, hemophilia A alleles would have had essentially these fitness properties). Given your answer for part **d**, what fraction of the alleles for the disorder will be lost each generation due to selection against affected males?

Since 1/3 of the X chromosomes are in males then 1/3 of the alleles will be in males and therefore will be lost.

(f) In order for the allele frequency for the disorder to remain at steady-state, new alleles must arise by mutation at the same rate as alleles are lost by selection. Given this constraint, what fraction of the alleles for the disorder would you expect to be new alleles and what fraction would you expect to be inherited from carrier females?

In order to maintain equilibrium, rate of loss must equal the mutation rate. From (e) we know that $\Delta q = 1/3\ q$ therefore $\mu = 1/3\ q$. This means 1/3 of the alleles are from new mutations.